National Oceanic and Atmospheric Administration Joint Hurricane Testbed Program

NOAA Award No. NA17OAR4590140

Title – Transition of Machine Learning Based Rapid Intensification Forecasts to Operations

Andrew Mercer, Principal Investigator
Associate Professor, Department of Geosciences, Mississippi State University
108 Hilbun Hall, PO Box 5448
Mississippi State, MS 39762-5448
mercer@hpc.msstate.edu
662-268-1032 x231

Brandy Akers, Organizational Representative
HPC[2] Director of Operations, Mississippi State University
Mailstop 9627
2 Research Blvd.
Starkville, MS 39759
bakers@hpc.msstate.edu
662-325-8814

Submission Date – January 30, 2019

Recipient Organization (Name and Address)

Project/Grant Period (July 1, 2017 – June 30, 2019)

Valid for Period ending January 30, 2019

Report Term or Frequency (semi-annual)

Final Annual Report? No

1. ACCOMPLISHMENTS

What were the major proposed ***goals, objectives, and tasks*** *of this project, and what was accomplished this period under each task? (a table of planned vs. actuals is recommended as a function of each task identified in the funded proposal)*

Table 1 below shows our objectives and scheduled goals. We have summarized the planned goals and our actual accomplishments during this reporting period.

Proposed Outcome	Anticipated Completion Date	Actual Outcome
Begin participation in 2018 Hurricane Testbed experiments	30-Nov-2018	Our ensemble was a member of the 2018 Testbed. We disseminated output in real-time to forecasters via the Automated Tropical Cyclone Forecast (ATCF) system
Conduct AI-based model verification at the conclusion of the Testbed, compare with NHC performance	30-Jan-2019	This work has been done for the current ensemble (more on this below) and the results to date are provided in Fig. 1.
Disseminate results at the AMS Annual Meeting	30-Jan-2019	Updates to the AI ensemble incorporating GFS Final Analysis (FNL) data were presented at the 2019 AMS annual meeting as part of the AI conference. Results are provided below.

As evidenced by the table above, expected completion dates for the third reporting period's tasks are in line with current project progress. We have no concerns that we will not be able to accomplish all tasks as outlined in our original proposal.

Are the proposed project tasks ***on schedule****? What is the cumulative percent toward completion of each task and the due dates? (table recommended)*

Table 2: Proposed remaining tasks from JHT proposal document with associated completion dates (from the initial proposal).

Proposed Task	Anticipated Completion Date	Percentage Completed
Port R-based AI codes to Python scripts	28-Feb-18	This work is ongoing due to updates to the AI ensemble described at the AMS annual meeting. (50%)
Begin Participation in 2018 Joint Hurricane Testbed	30 Nov-18	We have already participated in the 2018 Testbed. We intend to apply for a no-cost extension and participate in the 2019 Testbed as well with our updated model (100%)
Conduct AI-based model verification at conclusion of the Testbed, compare with NHC performance	30-Jan-19	Verification results are completed (see Fig. 1 below). Updated verification results for 2017 and 2018 are being developed presently for the new AI ensemble and will be presented at the Interdepartmental Hurricane Forum. (75%)
Publish results	30 Jun-19	One publication has already been accepted and a second is in development, to be submitted during this fourth period of the project. (50%)

We feel the work is on track for an on-time completion, despite some delays in the Python code porting. FNL fields are now fully implemented in the R-based AI ensemble, which has changed which members, as well as which features are part of the ensemble. These changes are yet to be implemented in the Python port of the R-based code. Further, the 2017 and 2018 seasons are not yet verified using the new FNL-based model. The government shutdown has limited our ability to acquire archived operational GFS forecasts for testing, but we will do so once the shutdown ends.

What were the major completed ***milestones*** *this period, and how do they compare to your proposed milestones? (planned vs. actuals table recommended)*

- The major planned milestones for this reporting period fell into two primary areas:

1. Participate in 2018 Joint Hurricane Testbed experiments – our initial AI ensemble model, which exclusively utilized SHIPS-RII predictors combined with storm characteristics derived from NHC best-track data, was set to update in real time during the 2018 Atlantic Hurricane Season. After feedback from NOAA personnel, we updated the format of the

output for our model and created readme documentation on this format to disseminate to forecasters and those interested in using our product. We feel we fully accomplished this task by our participation in the testbed and, other than a few technical challenges making our output available to NHC in real time, encountered no major issues meeting this milestone.

2. Conduct AI-based model verification at the conclusion of the Testbed and compare with NHC performance – While we are still waiting for 2018 best track data, operational NHC tracks allowed for preliminary verification statistics to be generated for all 2018 Atlantic TCs (Fig. 1).

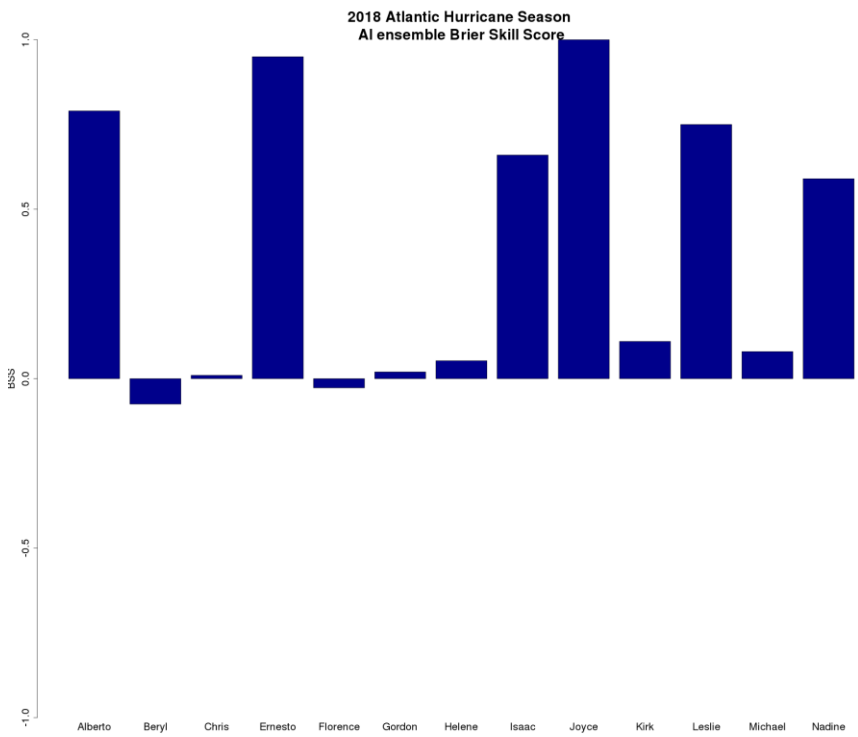

Fig. 1. Brier skill score (BSS) performance statistics for the 2018 hurricane season based on the outcomes of the 2018 Testbed experiments.

These results showed high skill for many storms and some negative skill in two events. Overall, our global results (BSS=0.04) were in line with the SHIPS-RII logistic model (BSS=0.04) and outperformed the original SHIPS-RII (BSS=0) and Bayesian models (BSS=0) for Atlantic TCs. These results were encouraging, though the AI ensemble did not outperform the SHIPS-RII consensus forecasts. Our intention is to reevaluate these statistics using the updated FNL-based AI ensemble (which was not included in the 2018 Testbed) to assess any benefits offered by this updated approach. These results will be presented at the TCORF and in the fourth reporting period's progress report.

In addition to the two proposed research goals, an updated AI ensemble based on FNL-derived predictors was developed and presented at the 2019 AMS annual meeting. In this model, a kernel principal component analysis (KPCA) was used to isolate important vertical layers of

FNL data useful in discrimination of RI/non-RI events and yielded considerable skill improvement over current forecast models. KPCA was selected as is useful for feature reduction in situations with large numbers of predictors and uses the nonlinear kernel matrix from support vector machines as its underlying similarity matrix. This in turn offers better clustering and separability that ultimately should improve classification in the AI ensemble. In the formation of the new AI ensemble, the following 43 predictors were retained:

- 8 SHIPS predictors, including -30°C brightness temperature, low-level relative humidity, mean 200 mb temperature, ocean heat content, average surface pressure, 850 mb – 250 mb shear magnitude, brightness temperature standard deviation, and Reynolds SST
- 4 NHC-derived predictors, including maximum wind speed, RI occurrence flag (1 if RI occurred previously, 0 otherwise), counts of previous RI timesteps for the given tropical cyclone, and previous 24-hour intensity change.
- 31 FNL-derived predictors, kernel PCs applied to 8 different vertical levels and variables:
    - 200 mb meridional wind (13 KPCs retained)
    - 850 mb specific humidity (3 KPCs retained)
    - 200 mb temperature (2 KPCs retained)
    - 200 mb equivalent potential temperature (2 KPCs retained)
    - 500 mb equivalent potential temperature (1 KPC retained)
    - 600 mb equivalent potential temperature (2 KPCs retained)
    - 500 mb vertical velocity (7 KPCs retained)
    - 600 mb static stability (1 KPC retained)

We noted that many of the FNL-derived fields had similarities with SHIPS-RII predictors (e.g. 200 mb temperature appeared from the SHIPS and from the KPCA, as well as low-level humidity). We are currently investigating some of the physical aspects as to why these differences exist, but it is important to note that the KPCA offers unique variability structures derived from the kernel matrix that provide unique information to the AI ensemble relative to the mean values used in SHIPS-RII. That is, including both the SHIPS-RII predictors and the KPCA does not contribute considerable overlap owing to the statistical differences underlying each predictor.

Upon retaining these predictors, a leave-one out cross-validation was done on all 5409 tested timesteps spanning 1999-2016. The resulting BSS for the testing data ranged from 0.27 to 0.42 depending on the AI configuration (Table 1):

Member	**RI Cutoff**	**POD**	**FAR**	**Bias**	**HSS**	**BSS**
SVM (cost=1,γ=0.05)	0.21	0.57	0.39	0.94	0.57	0.34
SVM (cost=1,γ=0.075)	0.15	0.64	0.43	1.11	0.58	0.36
SVM (cost=1,γ=0.1)	0.16	0.63	0.42	1.10	0.58	0.35
SVM (cost=10,γ=0.025)	0.22	0.67	0.40	1.12	0.61	0.41
SVM (cost=10,γ=0.05)	0.20	0.69	0.42	1.18	0.60	0.42
SVM (cost=10,γ=0.075)	0.19	0.67	0.41	1.14	0.61	0.40
SVM (cost=10,γ=0.1)	0.17	0.65	0.42	1.13	0.59	0.37
SVM (cost=10,γ=0.125)	0.26	0.57	0.38	0.92	0.57	0.35
SVM (cost=10,γ=0.15)	0.21	0.57	0.40	0.96	0.56	0.32
SVM (cost=10,γ=0.175)	0.14	0.60	0.43	1.06	0.56	0.29
SVM (cost=100,γ=0.025)	0.24	0.65	0.42	1.11	0.59	0.39
SVM (cost=100,γ=0.05)	0.19	0.70	0.43	1.22	0.60	0.42
SVM (cost=100,γ=0.075)	0.20	0.67	0.40	1.12	0.61	0.40
SVM (cost=100,γ=0.1)	0.17	0.66	0.42	1.14	0.60	0.37
SVM (cost=100,γ=0.125)	0.13	0.67	0.46	1.24	0.57	0.35
SVM (cost=100,γ=0.15)	0.16	0.61	0.42	1.06	0.57	0.32
Global Ensemble	Simple Mean: BSS = 0.398			Weighted Mean: BSS = 0.399		

These results show dramatic improvement over current upper limit of the SHIPS-RII skill scores (all members, including consensus) of roughly 0.2 and we intend to test this new ensemble on 2017-2018 GFS data (as independent testing) to verify the results hold up during new seasons.

What opportunities for training and professional development has the project provided?

- The PI attended the Complex and Adaptive Systems conference in Chicago, IL, as well as the AMS Annual Meeting in Phoenix, AZ, to present work related to this project (references are provided below).

How were the results disseminated to communities of interest?

- The PI presented work related to this project at the Complex and Adaptive Systems conference in Chicago, IL, and the AMS Annual Meeting in Phoenix, AZ. Work was also disseminated through published peer-reviewed conference proceedings at the Complex and Adaptive Systems conference and through communication with NOAA personnel.

What do you plan to do during the next reporting period to accomplish the goals and objectives?

- We plan to update the new FNL-based AI ensemble results for 2017 and 2018 Atlantic hurricane season data to assess its performance relative to current operational models.

- We will be submitting a manuscript describing the details of the FNL-based AI ensemble and its resulting performance relative to current NHC statistical models (SHIPS-RII and its constituents).
- We will attend the TCORF meeting in March 2019 to present the updated results from the AI ensemble and meet with NOAA personnel to discuss implementation of the model in the 2019 Testbed.
- We will pursue a no cost extension to allow us to participate in the 2019 Testbed experiments.

2. PRODUCTS

What were the major completed ***products or deliverables*** *this period, and how do they compare to your proposed deliverables? (planned vs. actuals table recommended)*

A previously unstated goal of this reporting period was to develop an updated AI ensemble using FNL observations, which was done successfully (as seen above). We also have begun converting this ensemble into Python for eventual direct use by the NHC. These are the primary deliverables to date, though we also have obtained validation statistics for our SHIPS-RII based AI ensemble used during the 2018 Testbed.

What has the project produced?

<u>Publications:</u>

Mercer, A., A. Grimes, and K. Wood, 2018: Multidimensional Kernel Principal Component Analysis of False Alarms of Rapidly Intensifying Atlantic Tropical Cyclones. *Procedia Comp. Sci.*, **140**, 359-366.

<u>Presentations:</u>

Mercer, A., A. Grimes, and K. Wood, 2018: Multi-Dimensional Kernel Principal Component Analysis of False Alarms of Rapidly Intensifying Tropical Cyclones. *Complex and Adaptive Systems Conference*, Chicago, IL, November 5, 2018

Mercer, A., A. Grimes, and K. Wood, 2019: An Updated Machine-Learning Ensemble for Atlantic Tropical Cyclone Rapid Intensification Forecasting. *18th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, AMS Annual Meeting*, Phoenix, AZ, January 7, 2019.

<u>Other Activities:</u>

Products available at arashi.geosci.msstate.edu/jht, including real-time AI forecast output and user guides for interpreting the output of the ensemble.

3. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on this project?
Andrew Mercer – Principal Investigator
Kimberly Wood – co-Principal Investigator
Alexandria Grimes – Ph.D. candidate

We have also had NOAA points of contact that we have interacted with:
Chris Landsea – lead point of contact, research to operations point of contact
Mark DeMaria – testbed point of contact
Erik Blake – testbed point of contact
Stacy Stewart – testbed point of contact
Jose Salazar – testbed technological point of contact

There have been no personnel changes during this reporting period.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

As this work spans the area of tropical meteorology and machine learning, we feel that the updates to our AI ensemble have presented more insight into the physical nature of RI within Atlantic TCs (based on resulting features) and improved methodologies for classifying unbalanced categorical data within machine learning. Specifically, the use of KPCA with SVMs is not commonly done in machine learning work but provided tremendous benefit over all previous efforts at ascertaining improved features and RI classification skill. This improvement has been attributed to the similar separability in KPCA and the SVM kernel matrix used in the AI ensemble. Additionally, the important retained fields suggest which physical features within the TC provide the greatest distinctions between RI and non-RI timesteps, and these features may be further explored in future work to diagnose physically what mechanisms are isolating RI from non-RI. Finally, the novel approach of identifying unique structures associated with RI false alarm forecasts has helped inform which features should be tested for RI classification, and this work has helped reduce the false alarm rate we were observing in the 2018 Testbed experiments. While we have not yet tested the new ensemble on 2018 data we anticipate the improvements we saw with the initial development of the new ensemble will be maintained for 2017 and 2018 TCs.

What was the impact on other disciplines?

There has been no direct impact on other disciplines as this product has not yet been adopted.

What was the impact on the development of human resources?

There has been no impact on the development of human resources.

What was the impact on teaching and educational experiences?

This project has been funding and training one Ph.D. candidate, and we recently began incorporating two undergraduate research students conducting side projects related to TC intensification.

What was the impact on physical, institutional, and information resources that form infrastructure?

There has been no notable impact as of this progress report. The closest to impact was the implementation of our AI ensemble in the 2018 Testbed and further development that will allow it to be included in the 2019 Testbed.

What was the impact on technology transfer?

There was no impact on technology transfer at this stage in the project.

What was the impact on society beyond science and technology?

As of this report no impact has been provided as the work is still ongoing. Once finished, the potential improved predictability of RI within Atlantic tropical cyclones will assist infrastructure preparations prior to hurricane landfall and provide forecasters much needed intensification guidance for tropical cyclones. The results could potentially save lives by facilitating earlier evacuations (as needed) and enabling earlier preparations for landfalling major hurricanes.

What percentage of the award's budget was spent in a foreign country(ies)?

No funds were spent in foreign countries.

5. CHANGES/PROBLEMS

As stated in the second reporting period's progress report, a major shift in the AI ensemble was needed after feedback from the 2018 TCORF meeting. Specifically, we revamped the AI ensemble using FNL fields instead of GEFS-reforecast fields as predictor inputs to determine if forecast improvements were possible. As seen in Fig. 2, this implementation led to an updated ensemble with fewer members that provided a significant boost in forecast skill relative to our previous implementation. We are currently finalizing the results of this new ensemble to make it available for the 2019 Testbed experiments.

6. SPECIAL REPORTING REQUIREMENTS

As this is a Joint Hurricane Testbed project, there are a few special reporting requirements. Outcomes for those requirements are provided after their questions below.

What is your assessment of the project's Readiness Level?

We feel that at the conclusion of the 2018 Testbed that some additional development is needed for our project. We therefore still find our project to be at a readiness level of 5. We intend to participate in the 2019 Testbed as well and expect the project will be at a level of 6 at the conclusion.

Project Test Plan

The project test plan has been provided in a previous progress report. No updates to this plan have been developed.

Transition to operations activities in the last 6 months

There have been no official transition to operations activities in the last 6 months. After discussions with NOAA personnel, we feel it best to allow our model to participate in the 2019 Testbed to demonstrate its promise for transition at the conclusion of the 2019 hurricane season.

Summary of testbed-related collaborations, activities, and outcomes

We have worked closely with our NOAA POC to revise the Research to Operations plan and our Testbed plan. We have also interacted with NOAA personnel at the Interdepartmental Hurricane Research Forum where we gained important advice to improve our models prior to the testbed. Many of those improvements have been implemented as a result of our collaboration with NOAA, with the rest being worked on presently.

Has the project been approved for testbed testing yet?

The project was already a participant in the 2018 Testbed. We will be using it in the 2019 Testbed as well.

What was transitioned to NOAA?

Nothing has been formally transitioned to NOAA as of this progress report. Indirectly, the ensemble code we have created has been writing e-deck formatted output that has been included as part of the ATCF. More formal inclusion in the ATCF will be done prior to the start of the 2019 Testbed experiment.

7. BUDGETARY INFORMATION

No major budgetary issues have arisen at this stage. The only budgeted funds to be spent by the end of the first quarter of the project were salary and benefits to the Ph.D. student, as well as her graduate tuition waiver. Budgeted travel to the AMS annual meeting and the Interdepartmental Hurricane Forum (for 2018) has been paid and is on track. The budget remains on target and we anticipate no budgetary issues going forward.

8. PROJECT OUTCOMES

What are the outcomes of the award?

No formal outcomes have been attained yet. Once the project finishes, the resulting AI ensemble should produce an outcome of improved RI forecasts, which should provide tremendous benefit to society in terms of early warning times for major hurricane landfalls and associated infrastructure preparations.

Are performance measures defined in the proposal being achieved and to what extent?

The objective at this stage of the proposed work was to obtain performance metrics for our AI ensemble in line with or exceeding NHC-derived RI models (specifically the SHIPS-RII). For the 2018 season, our AI ensemble performed consistently with the SHIPS-RII model, the Bayesian model, and the logistic model, outperforming two of the three, though it performed worse than the consensus model. We feel we are achieving but not exceeding our performance objectives as set forth in the original proposal.

However, the updated FNL ensemble yields considerably better results for the 1999-2016 data (using the leave-one-out approach, which should be methodologically similar to predicting on a new timestep). BSS values in this new ensemble now exceed 0.3 in most instances, and some ensemble members exceed 0.4 when tested on 5409 testing points. These results far exceed the baseline values seen in the SHIPS-RII literature. If similar performance is achieved when tested on 2017-2018, they may provide a vast improvement on currently available models. We will therefore have more conclusive results regarding performance metrics at the conclusion of the project.